

Exhibit 5

**UNITED STATES DISTRICT COURT
FOR THE SOUTHERN DISTRICT OF NEW YORK**

**NATIONAL ASSOCIATION FOR THE
ADVANCEMENT OF COLORED PEOPLE,
SPRING VALLEY BRANCH, *et al.*,**

Plaintiffs,

v.

**EAST RAMAPO CENTRAL SCHOOL
DISTRICT, *et al.*,**

Defendants.

ECF CASE

Case No. 7:17-cv-08943

**DISTRICT JUDGE
CATHY SEIBEL**

**MAGISTRATE JUDGE
JUDITH C. MCCARTHY**

DECLARATION OF DR. PETER A. MORRISON

1. My name is Peter A. Morrison.
2. I have been retained by the Defendant the East Ramapo Central School District (“the District”) to provide expert analysis regarding demographics in the District and to evaluate the opinions of Plaintiffs’ experts, Dr. Loren Collingwood and Dr. Matthew Barreto, and in particular Plaintiffs’ experts’ use of “Bayesian Improved Surname Geocoding” in their analysis. For my work on this report I was compensated at a rate of \$250/hour.
3. I am the President of Peter A. Morrison & Associates, Inc. I have 50 years of experience in applying demographic analysis to issues of public policy and law. For 35 years, prior to my retirement from The RAND Corporation, I served as RAND’s Senior Staff Demographer. I am the Founding Director of RAND’s Population Research Center. From 1989-1995, I served on the U.S. Census Bureau’s Advisory Committee on Population Statistics and thereafter as an invited participant in the Census Bureau’s

Working Group on 2010 Race and Ethnicity. Prior to my employment at RAND, I was an Assistant Professor at the University of Pennsylvania and a Research Associate at its Population Studies Center.

4. My principal expertise centers on evaluating proposed election districts and redistricting plans. I have made invited presentations on demographic aspects of redistricting before members and/or staff of the U. S. House of Representatives Subcommittee on the Census, the County Counsels' Association of California, the League of California Cities, the National League of Cities, and the Population Association of America.
5. I have authored/co-authored many peer-reviewed publications on methods for evaluating election districts and redistricting plans. These publications are widely used for instructional purposes in graduate courses in applied demography. Examples are: "From Legal Theory to Practical Application: A How-To for Performing Vote Dilution Analyses," *Social Science Quarterly* (2017); "Demographic Influences on Latinos' Political Empowerment: Comparative Local Illustrations," *Population Research and Policy Review* (1998); "Demographic Foundations of Political Empowerment in Multiminority Cities," *Demography* (1995); "Empowered or Disadvantaged? Applications of Demographic Analysis to Political Redistricting," in *Demographics: A Casebook for Business and Government*, Westview Press (1994); The Demographer's Role in the Local Boundary-Drawing Process, RAND, P-7711 (1991).
6. I have constructed and/or evaluated proposed congressional and legislative redistricting plans in California, Connecticut, Colorado, Florida, Georgia, Illinois, Maryland, Massachusetts, Mississippi, New York, Washington, and Wisconsin. I have testified

multiple times in state and federal court regarding my demographic analysis of the effects of congressional and legislative redistricting plans.

7. My *curriculum vitae*, which includes a list of my prior expert witness testimony, is appended to this declaration as Appendix A. Details of cases where I have provided testimony over the last four years are attached as Appendix B. I have no financial interest in the outcome of this litigation.
8. In preparing this declaration, in addition to the materials cited herein, I have reviewed the expert reports of Drs. Collingwood and Barreto that were disclosed in this case, the report of the District's expert, Dr. John Alford, and the materials associated with and disclosed with those expert reports. I have also reviewed Dr. Barreto's deposition transcript, the report of Plaintiffs' expert demographer Dr. William Cooper, and certain materials disclosed with his report. Finally, I have drawn upon data from the 2010 decennial census and the annual American Community Survey.

Development of BSG and BISG

9. For many years, statisticians, demographers, political scientists, and others have confronted a persistent problem: inferring the race/ethnicity of an individual or group of individuals whose self-identified race/ethnicity is unknown or unavailable.
10. In 2008, my co-authors and I advanced a new approach to solving this problem, called Bayesian Surname Geocoding ("BSG"). BSG uses Bayes' theorem to estimate an individual's race/ethnicity based upon the person's surname and residence address.¹

¹ Marc N. Elliott, Allen Fremont, Peter A. Morrison, Philip Pantoja, & Nicole Lurie, *A New Method for Estimating Race/Ethnicity and Associated Disparities Where Administrative Records Lack Self-Reported Race/Ethnicity*, Health Research Service (2008), <https://www.ncbi.nlm.nih.gov/pubmed/18479410>.

11. In 2009, we published a further refined version of our methodology, called Bayesian Improved Surname Geocoding (“BISG”). BISG expands upon the initial BSG model that we created by incorporating additional surname information from the US Census Bureau and other scientific sources.²
12. In short, BISG uses two sources of data to estimate the probability that a person’s race/ethnicity is White, Black, Asian and/or Hispanic: census data characterizing the racial/ethnic makeup of the immediate neighborhood where that person lives; and several scientifically validated surname lists with a proven ability to distinguish a person’s probable racial/ethnic identity given that person’s last name. The strength of BISG is its integration of these two complementary data sets to furnish an estimated probability that someone belongs to one or another of six categories: Asian, Hispanic, Black, White, multiracial, or American Indian/Alaska Native.
13. BISG performs best when applied to populations that live in racially homogenous neighborhoods and whose names generally comport with national naming conventions. Setting aside the statistical details, the logic of BISG rests upon a few basic ideas. For example, a voter whose last name is Rodriguez and who lives in a mostly-Hispanic neighborhood has a strong likelihood of self-identifying as “Hispanic” on the decennial Census. The US Census Bureau’s internal research has documented that likelihood among 5,609,592 persons.³ By contrast, a voter whose last name is Smith, living in a

² Marc N. Elliott, Peter A. Morrison, Allen Fremont, Daniel F. McCaffrey, Philip Pantoja, & Nicole Lurie, *Using the Census Bureau’s Surname List to Improve Estimates of Race/Ethnicity and Associated Disparities*, Health Services & Outcome Research Methodology (2009), <https://link.springer.com/article/10.1007/s10742-009-0047-1>.

³ Reference here is to the Passel-Word list of 5,609,592 matchable Spanish Origin Research person records, of which 597,533 (10.7 percent) reflect a self-reported Hispanic origin. See R. Colby Perkins, *Evaluating the Passel-Word Spanish Surname List*, US Census Bureau, Population Division Working Paper No. 4, August 1993, <https://www.census.gov/population/www/documentation/twps0004.html>.

mostly-non-Hispanic neighborhood, has a minimal likelihood of self-identifying as “Hispanic” on the decennial Census (with rare exceptions, such as a married voter registered as “Maria [Rodriguez] Smith,” who may self-identify as “Hispanic”).

14. By the same logic, surname analysis alone works well for Asians, who also have distinctive last names.⁴ Here, too, a voter’s neighborhood demographic profile may figure in. The surname Lee may indicate that a person is likely to self-identify as “Asian” in a predominantly Asian neighborhood but not necessarily in most other neighborhoods.
15. For persons who self-identify as “Black,” surname analysis has proven to be of very limited use, since few surnames are unique to Blacks. Here, however, the neighborhood where a voter resides is useful for gauging the probability of that voter self-identifying as “Black.” In short, heavily Black neighborhoods are a more useful basis for distinguishing between Blacks and Whites. However, the limitation here is that not every Black voter lives in a heavily-Black neighborhood.
16. Under certain circumstances, and for addressing particular needs, BISG provides valuable information about the racial composition of a group. BISG evolved out of the needs of health care insurers and providers to estimate the racial and ethnic composition of the populations they serve. In many instances, that information is totally unavailable.
17. Like any statistical tool, BISG has recognized limitations. Despite those limitations, BISG can yield reliable estimates of the racial and ethnic composition of populations

⁴ See Diane S. Lauderdale & Bert Kestenbaum, *Asian American Ethnic Identification by Surname*, Population Research and Policy Review (2000), <https://link.springer.com/article/10.1023/A:1026582308352>.

where such information has not been recorded at all or cannot legally be made available owing to strict confidentiality laws.

18. For many years, demographers such as myself have used official data published by the Census Bureau to estimate the racial/ethnic composition of eligible voters (i.e., voting-age citizens). Today, such estimates are routinely made for election districts, precincts, polling places, and the like, based upon American Community Survey (ACS) data. Political scientists have devised and relied upon ever more sophisticated methods to translate such hyperlocal data into credible estimates of legally recognized standards upon which courts now rely. The table below shows each racial/ethnic group's relative share of the District's total population and those who are eligible voters (the citizen voting-age population, or CVAP).

Total Population and Citizen Voting-Age Population, 2013-17: East Ramapo Central School District				
	Total Population		CVAP	
	No.	%	No.	%
All residents	121,835	100%	61,123	100%
White Non-Hispanic	72,680	60%	37,454	61%
Black alone	25,110	21%	15,337	25%
Asian alone	4,253	3%	2,978	5%
Hispanic	18,866	15%	4,798	8%
Source: US Census Bureau, 2013-17 ACS 5-year estimates, Tables B05003, B05003B, B05003D, B05003H, B05003I.				

19. As seen above, the proportion of Black and Asian residents eligible to vote is proportionally higher than each group's share of all District residents (Blacks: 25% vs.

21%; Asians: 5% vs. 3%). Hispanic residents differ. They comprise only 8% of residents eligible to vote despite being 15% of all District residents. Corresponding CVAP measures for each of the Districts 10 precincts would reflect most directly the *local* political presence of each group as eligible voters within that specific area.

20. BISG's estimates, by contrast, necessarily derive from *nationally-representative* surname data (used in conjunction with measures of local neighborhood population composition). Herein lie a number of hidden unknowns which persuade me that BISG ought not be the preferred choice of Plaintiffs in this case. Instead, I favor adhering to the well-established methods that political scientists now use to derive credible estimates of legally recognized standards upon which courts must rely. Simply put, this is an instance of favoring the devil we know, not the devil(s) we don't know.
21. As a coauthor of efforts to develop, refine, and evaluate BISG, I am troubled by the real prospect of one or several distorting influences sure to go unrecognized in Plaintiffs' application of BISG. I based this opinion upon (1) my considerable first-hand experience in applying Spanish and Asian surname lists in local contexts, (2) my detailed familiarity with the US Census Bureau's technical evaluations of the Bureau's List of Spanish Surnames (on which BISG relies), and (3) the *unknown* potential risks associated with using BISG, given obvious differences between the citizen voting-age population of the District, the US as a whole, and the State of Florida in particular (which was used in a validation study on which Plaintiffs' experts rely). No one—myself included—knows whether or how those differences might inadvertently misinform an expert on either side in this case.

22. Drawing upon my considerable first-hand experience, I recognize that BISG's surname component embodies the inherent limitations of a nationally-representative data set. Those limitations are inevitable where valid nationally-representative data are applied to a particular place or neighborhood. We must assume that what has been measured nationally applies locally to the population of interest—here, each of the District's 10 election precincts. If this assumption does not hold, then BISG's estimates of race might well mislead any expert. There are ample grounds for questioning this assumption here.⁵
23. Next, I draw upon my detailed familiarity with the US Census Bureau's technical evaluations of the Bureau's "List of Spanish Surnames" on which BISG relies. The "List of Spanish Surnames" identifies persons of *probable* Hispanic origin. The List is unique in that it has undergone thorough scientific evaluation. Its detection characteristics are well understood, and its limitations have been carefully documented. The most exhaustive evaluation furnishes researchers two important statistics for judging the List's effectiveness: (1) the surname commission (SCOM) rate ("false positives"), i.e., the percentage of people whose surnames appear on the List but who reported as "Non-Hispanic" on the 1990 Decennial Census Spanish Origin Research File; and (2) the surname omission (SOM) rate ("false negatives"), i.e., the percentage of people who reported "Hispanic" but whose surname does not appear on the List.
24. The lower the SCOM "false positive" rate, the more reliable the surnames on the List are at detecting the Hispanic population; the lower the SOM "false negative" rate, the greater

⁵ The use of "Spanish-surname" registration can be problematic. At least one district court has noted the problems associated with "Spanish-surname analysis" because of its tendency to misidentify Hispanic persons as non-Hispanic and vice-versa. *See United States v. Alamosa County*, 306 F. Supp. 2d 1016, 1022 (D. Colo. 2004). That court held that the expert testimony based on Spanish-surname data, while probative, should be afforded reduced weight, and noted that self-identification data provides a more reliable means of determining ethnicity. *Id.*

the proportion of the Hispanic population the List detects. Ideally, both would be zero.

In fact, they vary significantly by region, as shown in the Table below:

Variation in Spanish Surname List Error Rates Across the United States

<u>Geographic Region</u>	<u>SURNAME COMMISSION</u>		<u>SURNAME OMISSION</u>	
	<u>Rate</u>	<u>Std Error</u>	<u>Rate</u>	<u>Std Error</u>
Mexican Southwest	7.11%	0.05%	16.05%	0.06%
Puerto Rican Northeast	8.26%	0.08%	20.86%	0.11%
Florida	10.15%	0.18%	26.06%	0.24%
Hawaii	61.39%	0.79%	59.54%	0.82%
Rest of United States	26.74%	0.19%	38.57%	0.19%

Source: Table 4 in R. Colby Perkins, *Evaluating the Passel-Word Spanish Surname List: 1990 Decennial Census Post Enumeration Survey Results*. US Census Bureau, Population Division, Technical Working Paper No. 4 (1993).

25. Such sizable variations in rates of Hispanic misidentification are a definite cause for concern. Translating national surname-based estimates of race or ethnicity into hyperlocal estimates of a particular group's share of all voters in a tiny election precinct is fraught with uncertainties. Doing so substitutes one uncertainty (the race/ethnicity of actual voters) for another uncertainty (the turnout rate among eligible voters whose race/ethnicity is known).
26. Lastly, I emphasize the unknown risks of adopting a new method like BISG in the face of likely differences between the citizen voting-age population of the District and that of the entire US or just one state. The software that Plaintiffs utilized to execute their own preliminary analysis was validated based upon a single validation study confined just to just one state.
27. In paragraph 18 of Barreto and Collingwood's Expert Report, Plaintiffs' experts state: "To assess the racial composition of the East Ramapo electorate with the **greatest feasible accuracy**, we also applied BISG to the voter file for the 2017 election using a

statistical package for voter analysis developed by Imai and Khanna, the authors of a peer-reviewed article on this topic published in a leading journal on political science methodology.” [Emphasis added]

28. I have reviewed the referenced article. The article’s conclusions are based exclusively on a validation experiment conducted in only one state. Surely, a single state cannot be regarded as nationally representative, which is why I dispute the above-quoted statement (specifically “greatest feasible accuracy”). Plaintiffs favor BISG over established methods that utilize American Community Survey measures of the actual eligible voters in specific precincts. I regard that decision as perilous and unjustified in light of what is known.
29. Validation on just one state introduces the most troubling of all unknowns--those we do not recognize (“the devil one doesn’t know”). Consider the following two hypothetical examples. Suppose that a national surname list shows that 80% of people named “*Germaine*” are White (since the surname is French, and most French people are White). However, the surname “*Germaine*” could mislead us were it used to estimate race in East Ramapo, which happens to have a sizable Haitian population. Haitians are predominantly Black and often have French names. Furthermore, East Ramapo also has a population of Sephardic Jews, who sometimes have such surnames as “Laredo,” “Toledano,” “Rodriguez,” and “Moreno”--all of which appear on the Census Bureau’s “List of Spanish Surnames” and hence would be classified as “Hispanic.”
30. BISG assumes that one’s surname is independent of one’s location. This assumption does not necessarily hold in small communities or those where extended families predominate within the same neighborhood. In such settings, a voter’s address provides

meaningful information regarding one's surname, yet incorporating both pieces of information may place too much (or too little) weight on the race information derived from an address.

31. This assumption is likely violated in the District. It is my understanding that the District is home to a large Orthodox Jewish population, in which extended families tend to live together in concentrated areas.
32. Like any estimation method, BISG works much better when it is applied to large data sets and when it is used to predict race in the aggregate. At the aggregate level, each individual error has a negligible effect on the overall result. When my coauthors and I first introduced BISG, it had been tested on a national data set of several million people.

Evaluation of Barreto and Collingwood's Use of BISG

33. In their rebuttal Report, Dr. Collingwood and Dr. Barreto quote my coauthored 2008 publication to support the categorical claim that "BISG is the best available method for estimating individuals' race/ethnicity."⁶ My co-authors and I made no such claim.
 - a. First, the 2008 publication upon which Dr. Collingwood and Dr. Barreto rely was about *BSG*, the initial methodology that inspired our subsequent refined *BISG* methodology, published thereafter. The 2008 article makes no reference to *BISG*. This distinction between the two methods was apparently lost on Dr. Barreto, who claimed at his deposition that *BISG* and *BSG* are the same.⁷

⁶ In paragraph 18 of their rebuttal report, they state: "Overall, the findings indicate BISG is the best available method for estimating individuals' race/ethnicity. For example, Elliot et al. (2008) find that BISG is '74 percent more efficient than geocoding alone...and 56 percent more efficient in estimating the prevalence of racial/ethnic groups, outperforming the non-Bayesian hybrid on both measures (p. 1722)."

⁷ See Deposition of Matthew Barreto ("Barreto Tr.") at 217-18.

- b. Second, the specific passage quoted by Collingwood and Barreto refers to statistical efficiency—a concept that is distinct from statistical accuracy.

Statistical efficiency refers to the amount of information that can be ascertained from a fixed quantity of data. Accuracy, on the other hand, refers to the extent to which a method generates correct estimates or predictions. Once again, this distinction apparently eluded Dr. Barreto who, at his deposition, suggested that efficiency is synonymous with accuracy.⁸

34. It is incorrect to state categorically that “BISG is the best available method for estimating individuals’ race/ethnicity.” The question of whether a particular method is “best” depends on many factors, including one’s objective and the quality of the underlying data. The best available method for assessing a person’s race/ethnicity derives from that person’s own self-report on the American Community Survey or the decennial census. On a Census Bureau questionnaire, if you check “Hispanic” you are Hispanic; otherwise, you are not Hispanic. Each person’s direct responses to questions about one’s Hispanic ethnicity and one’s race are reported as true measures of how that person self-identifies. Neither probabilities nor estimates enter in here.
35. In their report, Collingwood and Barreto purport to have used BISG to estimate the proportion of voters in each of the District’s ten voting precincts that are White, Black, or Hispanic. Based on my understanding as a participant in its development, I doubt that BISG would generate reliably accurate predictions of voters’ race/ethnicity in this setting. First, the demographics of the District are so unusual that one cannot safely assume that circumstances locally correspond to national patterns. Second, the absolute numbers of

⁸ Barreto Tr. 217-21.

voters District-wide is very small. Third, apart from several areas that are almost entirely White, the District's Black and Hispanic residents are distributed roughly evenly throughout the District, with little variation at the block level.⁹ This means that, for most District residents, one's immediate neighborhood location does little to improve an estimate of one's race/ethnicity.

36. The greatest barrier to using BISG to predict voters' race in District elections is the small number of voters that are involved. District elections typically draw around 14,000 voters, less than one percent of the sample size on which our original BISG model is based. To my knowledge, BISG has never been applied to a population as small as the District's, much less undergone a validation analysis based upon a population that small. BISG might well prove unreliable under such circumstances.
37. In their reports, Collingwood and Barreto use BISG estimates of voter race as an input into an EI analysis of voter preferences. However, they do not consider whether the District satisfies any of the assumptions required for the reliable use of BISG, nor did Barreto identify any assumptions when asked at his deposition.¹⁰
38. Additionally, Collingwood and Barreto did not disclose the actual results of their BISG analysis, so I cannot evaluate whether their results seem reasonable, either at the precinct level or viewed in the aggregate.
39. Collingwood and Barreto were not the only experts to produce estimates of voter choice by race. The District's political science expert, John Alford, produced EI estimates of voter choice from precinct-level CVAP data.

⁹ See Cooper Report at 16 Fig. 9.

¹⁰ Barreto Tr. 98-99.

40. Collingwood and Barreto adopt an extreme position. They contend that their BISG EI estimates are meaningfully different from Dr. Alford's CVAP estimates of voter choice, and that the CVAP EI estimates should be discarded because they were derived from data that was "less precise."¹¹ I do not agree with either conclusion.
41. First, the voter-choice estimates that these two methods yield are not altogether different, judging from the confidence intervals associated with both sets of estimates.
42. Second, I disagree with Collingwood and Barreto's assertion that the voter choice estimates generated by a BISG EI analysis are necessarily preferable or more accurate than those generated by a CVAP EI analysis. I have elaborated above upon the many uncertainties that may threaten the accuracy of BISG estimates in this particular District.
43. Collingwood and Barreto's vote choice estimates are especially worrisome because their estimation method did not take account of the uncertainty surrounding their BISG estimates. In other words, their estimates have two layers of uncertainty—the uncertainty associated with the ecological inference method, which is reflected in the confidence intervals that accompany their estimates, and the uncertainty surrounding their BISG inputs, which Collingwood and Barreto seem to ignore.¹²
44. I strongly favor the standard practice, which is to use CVAP data to estimate voter choice by race. Indeed, the Department of Justice makes CVAP data available annually for the specific purpose of analyzing actual and potential voting behavior. The fact that CVAP is the long-established standard here persuades me that Plaintiffs should conduct a CVAP analysis and give weight to the results, even if only as a control on their riskier alternative of using BISG.

¹¹ Barreto Tr. 159-60.

¹² Barreto Tr. 111.

45. I disagree with Collingwood and Barreto's assertion that a BISG analysis is better because it is based upon "more precise" data. The concern here is not precision but uncertainty. CVAP and BISG each include a measure of uncertainty. The distinctive strength of CVAP data is that they furnish direct and accurate measures of the race/ethnicity of eligible voters at hyperlocal levels of geography (e.g., individual precincts). Generating voter choice estimates using CVAP does necessitate using well-established methods like EI to estimate voter turnout.
46. A BISG analysis, on the other hand, provides accurate information about voter turnout because it measures just those eligible voters who actually turned out to vote. The drawback, however, is that BISG can only estimate the *odds* of a voter's race/ethnicity rather than actually knowing what each person has self-reported. Accordingly, I doubt that Barreto's and Collingwood's decision to use BISG data instead of CVAP data for their EI analysis results in greater precision—it merely substitutes one kind of uncertainty for another.
47. As explained above, when it comes to District elections, I doubt that BISG would generate reliable estimates. In that sense, the debate over BISG and CVAP in this particular case is easily resolved: Were there a large discrepancy between the BISG and CVAP estimates, I would rely upon the CVAP estimates.
48. Having viewed the results themselves, however, my opinion is that there is no discrepancy to resolve, as explained above. Not surprisingly, both EI analyses generated estimates with extremely wide confidence intervals for Black and Hispanic voters, based upon just 10 precincts. The confidence intervals for Black voters scarcely differ. Yet

one cannot help but be troubled by margins so thin that a mere handful of misclassified voters would yield a very different result.

Summary of Conclusions

49. Collingwood and Barreto's conclusions that BISG can be used to "assess the racial composition of the East Ramapo electorate with the greatest feasible accuracy" and that "BISG is the best available method for estimating individuals' race/ethnicity" are incorrect.
50. The small size of the District's voting population, in tandem with various features of the District's demographics, means that BISG is not likely to yield a reliable estimate of voter race in District elections. Under the circumstances, the preferred method would be to use CVAP as the data input for EI, because CVAP is collected specifically for this purpose and is uniquely capable of reflecting hyperlocal variations submerged in a nationally representative data set.
51. The large confidence intervals associated with the various voter choice estimates show that the voter choice estimates produced by Collingwood and Barreto's BISG analysis do not meaningfully differ from the estimates produced by Dr. Alford's CVAP analysis. However, if there were there a large discrepancy between the BISG and CVAP estimates, I would rely upon the CVAP estimates.

Pursuant to 28 U.S.C. § 1746, I declare under penalty of perjury that the foregoing is true and correct.

Executed on December 7, 2018



Peter A. Morrison